



HAL
open science

Une méthodologie sur corpus pour l'écriture en FOU

Cristelle Cavalla

► **To cite this version:**

Cristelle Cavalla. Une méthodologie sur corpus pour l'écriture en FOU. Points Communs - Recherche en didactique des langues sur objectif(s) spécifique(s), 2019, Démarche FOS et pratiques de classe : articulation entre ingénierie de formation et ingénierie pédagogique, 47, pp.91-108. hal-02132668

HAL Id: hal-02132668

<https://hal.science/hal-02132668>

Submitted on 17 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cavalla Cristelle, 2019, Une méthodologie sur corpus pour l'écriture en FOU. *Points Communs* 47, 91-108. URL : <https://www.lefrancaisdesaffaires.fr/professeurs/ressources/points-communs/>

Une méthodologie sur corpus pour l'écriture en FOU

Cristelle CAVALLA, Université Sorbonne Nouvelle-Paris 3, DILTEC EA2288

Cet article¹ a pour double objectif de présenter d'abord une approche didactique où le rôle de l'enseignant est celui d'un guide qui enseigne une méthodologie d'apprentissage et ensuite deux outils pour l'écriture académique : les corpus numériques et le lexique phraséologique transdisciplinaire.

L'hypothèse de départ est issue d'un constat après plusieurs années d'enseignement du FOU² auprès d'étudiants de Master et de Doctorat avec un niveau B2-C1 en français. Il s'agit de vérifier si la présentation d'une méthode d'apprentissage de l'écrit universitaire serait moins anxiogène qu'une entrée par la langue. En effet, les étudiants non spécialistes en langue, et devant rédiger en L2 (français ou autre) avec un niveau B2-C1, sont parfois contrariés par la langue de spécialité qu'ils doivent manipuler tandis qu'ils maîtrisent leur discipline (chimie, gestion, histoire...). Afin d'aider ces étudiants à rédiger, il est apparu qu'un cours de méthodologie de l'écrit universitaire tourné d'abord vers un outil numérique, puis vers un aspect spécifique de la langue, diminue leur anxiété face à la langue.

Ainsi, notre objectif global se structure en deux grandes parties : 1/aider les apprenants à utiliser les corpus numériques pour leurs écrits académiques et 2/les aider à découvrir le lexique transdisciplinaire approprié à ce type d'écrit. Pour ce faire, nous présentons une recherche-action développée sur plusieurs années afin d'avoir des données à la fois côté apprenants et côté enseignants ; puis nous aborderons les outils : les corpus numériques et le lexique transdisciplinaire ; enfin, nous donnerons les résultats constatés.

1. Mise en place d'un cadre didactique

¹ Cet article est la reprise partielle d'un article paru chez ISTE sous la direction de A. Tutin et M-P. Jacques (Cavalla 2018).

² Français sur Objectif Universitaire.

L'approche didactique qui nous intéresse ici est celle développée dans le cadre du Français sur Objectif Universitaire (FOU) (MangianteParpette, 2011) en lien étroit avec l'ingénierie de formation (Leclercq, 2003). Au plan didactique, nous avons pris en compte différentes étapes développées sur plusieurs années au fil de projets de recherche (ANR Scientext³ avec le projet FULS⁴ puis l'ANR TermITH⁵) et de suivis d'étudiants :

- L'analyse des besoins⁶ : enquête auprès de doctorants, en L2 pour le français, du domaine universitaire de Grenoble par Bard (Bard, 2007).
- La collecte de données authentiques : élaboration de corpus numériques d'écrits scientifiques de locuteurs natifs (ANR Scientext puis TermITH).
- Le traitement des données : analyses linguistiques sur la démarcation et le positionnement de l'auteur dans le cadre de plusieurs mémoires de Master FLE, éléments qui ont notamment servi aux thèses de Mroué (2014), Tran (2014), Hatier (2016) et Yan (2017).
- L'élaboration d'une séquence : d'abord en ligne avec le projet FULS (Cavalla, 2014; Cavalla, 2015), puis en présence à l'aide des corpus numériques dans le cadre des thèses de doctorat de Tran (2014) et Yan (2017).
- L'élaboration des activités didactiques : les activités développées sont en lien avec le corpus et prennent en compte la dimension « numérique » de l'outil.

Une telle organisation est représentée dans le schéma 1. L'article se concentre sur les étapes 5 et 6 du programme FOU, tandis que les étapes précédentes ont fait l'objet d'autres publications.

³ Responsables A. Tutin et F. Grossmann. Interface ScienQuest : <https://corpora.aiakide.net/scientext19/>

⁴ Formes et Usages des Lexiques de Spécialités. Responsable C. Cavalla : http://webtek-66.iut2.upmf-grenoble.fr/index.php?dossier_nav=689

⁵ Terminologie et Indexation de Textes en Sciences Humaines. Responsable E. Jacquey : <http://www.atilf.fr/ressources/termith/>

⁶ Ces travaux ont été réalisés en partie par des étudiants de Master FLE de Grenoble. Ces derniers ont participé au développement des projets Scientext, FULS et TermITH.

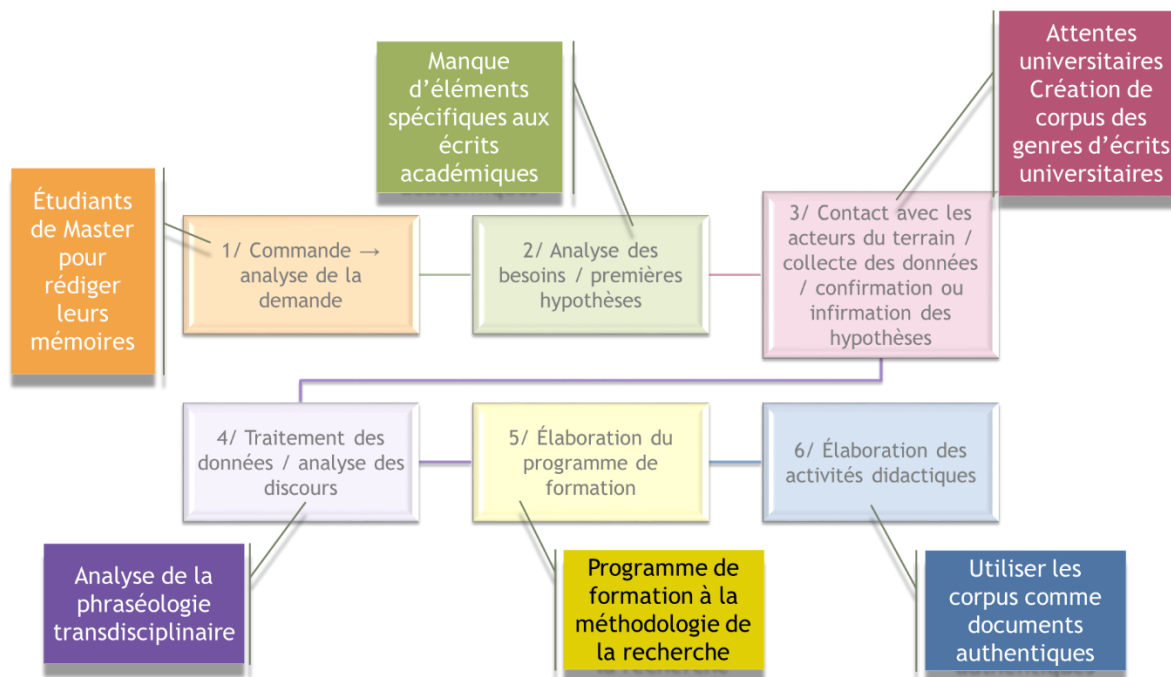


Schéma 1 : La démarche FOS/FOU pour des étudiants de Master et Doctorat

Les extraits d'activités choisis ci-dessous, intègrent les deux aspects évoqués : les corpus numériques et le lexique phraséologique transdisciplinaire.

2. Premier outils : les corpus numériques

Le corpus numérique utilisé en ligne est l'association de Scientext et TermITH. Deux interfaces ont été développées : ScienQuest et DiCorpus. ScienQuest permet une interrogation libre ou guidée (et avancée pour les plus aguerris) qui donne les résultats sous forme de concordanciers. L'utilisateur peut soit les exporter pour les trier et créer des exercices (côté enseignant), soit les utiliser en les parcourant soit en les comparant pour s'aider à l'écriture (côté apprenant). DiCorpus (TranFalaise, 2016) est orienté vers l'apprenant et l'aide à la rédaction scientifique. L'interface propose deux types de requêtes : l'une onomasiologique (par expression) et l'autre sémasiologique (par fonction discursive). L'utilisateur peut choisir une expression ou une fonction discursive présente dans les écrits scientifiques (par exemple « à propos de » (Copie d'Écran 1 ci-dessous) ou « enchaînement thématique (CE2)), pour la fonction discursive, la recherche se poursuit de la sorte : l'utilisateur choisit une sous-catégorie de cette fonction (par exemple : « insister sur une idée ») et des expressions appropriées (des éléments phraséologiques) apparaissent (dans ce cas : *à ce propos, en l'occurrence, à propos de...*).

Séquences lexicalisées à fonction discursive dans les écrits scientifiques

MODE D'ACCÈS	EXPRESSION	CATÉGORIE
<ul style="list-style-type: none"> ▶ Accès par sens ▶ Accès par expression 	à propos de (OK) d'emblée (**) tout d'abord (***) par exemple (****) en l'occurrence (**) à cet égard (*) à ce sujet (*) à propos de (****) en conclusion (*) pour conclure (*) par rapport à (****) il en est de même pour (*) conformément à (**)	adverbe POSITION DANS LA PHRASE en position postverbale finale PORTÉE - Locale (syntagme nominal) DÉFINITION là-dessus, sur ce sujet FRÉQUENCE ****

Copie d'écran 1 : Entrée onomasiologique de *à propos de* dans DiCorpus

MODE D'ACCÈS	POUR EXPRIMER...	SOUS-CATÉGORIE	EXPRESSION	CATÉGORIE
<ul style="list-style-type: none"> ▶ Accès par sens ▶ Accès par expression une exemplification ... un enchaînement thématique ... une énumération ... une reformulation ... une comparaison	... insister sur une idée ... ajouter une nouvelle information ... mettre en relief deux aspects	à ce propos (**) en l'occurrence (**) à cet égard (*) à ce sujet (*) à propos de (****)	adverbe POSITION DANS LA PHRASE en position postverbale finale PORTÉE - Locale (syntagme nominal) DÉFINITION là-dessus, sur ce sujet FRÉQUENCE ****

Copie d'écran 2 : Entrée sémasiologique de *à propos de* dans DiCorpus

Sur la copie d'écran 2 on distingue les fréquences des éléments dans le corpus ce qui donne un aperçu de leur utilisation dans le champ scientifique (ici on voit que *à propos de* est fréquent avec 4*).

La copie d'écran 1 donne à voir la position de l'élément dans la phrase (plutôt postverbale ou finale pour *à propos de*), sa définition sommaire, sa catégorie grammaticale et sa portée sémantique qui contribue à une insertion appropriée dans la phrase. Il manque des exemples, mais les développeurs sont en train d'y réfléchir soit en ajoutant un lien direct vers ScienQuest, soit en indiquant quelques phrases.

La séquence présentée ci-après tiendra compte de plusieurs facteurs cognitifs intégrant le DiCorpus : 1/ l'aide à la mémorisation en s'appuyant sur les acquis antérieurs des apprenants, 2/ l'évitement de la traduction sauf si deux éléments phraséologiques sont identiques en L1 et L2, 3/ l'appropriation simultanée de l'outil et des éléments linguistiques. Nous entrons par l'utilisation de l'outil (le corpus numérique) et intégrons ensuite l'élément linguistique. L'approche est celle du FOU avec une analyse des besoins effectuée en amont, la collecte des données est recensée dans les corpus et leur traitement a été réalisé au fil des recherches des membres des projets. Nous proposons les exercices prenant en compte l'appropriation du corpus par les apprenants et la découverte des éléments phraséologiques spécifiques à l'écriture académique.

La séquence se déroule en trois temps pédagogiques :

- la découverte de l’outil corpus,
- son appropriation d’abord libre, puis avec des contraintes linguistiques (recherche d’éléments phraséologiques),
- la réutilisation pour la présentation d’un sujet de mémoire avec un diaporama.

Cette séquence a été mise en place auprès d’étudiants de différents Masters en Sciences Humaines⁷ (langues, psychologie, droit) avec des L1 variées. Leur niveau en français variait du B2 au C1 et ils devaient rédiger un mémoire en français. Le cours s’est déroulé sur 20h en un semestre.

La séquence débute par un remue-méninge sur les connaissances et les représentations de l’écrit scientifique par les apprenants. Cet échange collectif permet d’ajuster des représentations erronées (« tous les chercheurs savent écrire dès le début de leur carrière ») et de conforter celles qui apparaissent plus appropriées (il faut s’entraîner à rédiger). Les objectifs sont présentés au premier cours dans l’ordre suivant :

- Découvrir et s’approprier l’utilisation des corpus numériques : trouver les éléments linguistiques qui seront utiles pour la rédaction finale ;
- Préciser les connaissances linguistiques sur les écrits académiques : le vocabulaire spécifique ;
- Réfléchir à une méthodologie de rédaction : penser à utiliser l’outil pour vérifier quels sont les éléments fréquents, pour lister/s’approprier les éléments linguistiques utiles à la rédaction...

2.1. Découverte et appropriation de l’outil

L’entrée est donc outillée. Cela a pour effet de rassurer les étudiants parfois stressés face à des éléments linguistiques inconnus. La démarche est dite « directe » (ChambersO’Sullivan, 2004), car nous plaçons les étudiants devant les ordinateurs et nous ne leur fournissons pas d’éléments présélectionnés pour deux raisons : leur niveau en langue leur permet d’opérer des tris et les choix de mots dépendent trop de leurs écrits personnels ; il nous serait alors difficile de deviner les besoins individuels des étudiants-apprenants.

⁷ Elle a été testée plusieurs fois auparavant avec des étudiants d’autres disciplines universitaires: chimie, gestion, histoire...

Le cours commence alors par ce qu'ils connaissent (le « déjà-là ») sous forme de vérification de leur « intuition linguistique ». La question posée (« Dit-on plus souvent *Je prends un risque* ou *J'attrape un risque* ? », « Dites-vous *Je mets la table* ou *Je mets le couvert* ? ») les conduit sur *Ngram Viewer*⁸, interface qui permet de vérifier la fréquence de lexies isolées ou d'éléments phraséologiques dans un corpus écrit. Souvent les étudiants connaissent cet outil fort apprécié pour la vérification de son intuition en langue étrangère. Il leur est alors demandé de trouver une traduction de l'élément fréquent appropriée à un contexte particulier de leur choix, grâce à *Linguee* (site connu des apprenants). Enfin, le dernier exercice de découverte permet de vérifier deux choses avec eux et pour eux : 1/ leur méthode de recherche en ligne, 2/ ce que le web ne nous donne pas à voir pour l'aide à l'écrit académique. La demande paraît simple *a priori*, et reste pourtant sans réponse pour certains étudiants : « Dans quel type de phrase trouve-t-on l'expression *en particulier* ? Comment s'en servir dans les écrits académiques ? ». Dans chacune des trois formations dispensées à des apprenants (Master et Doctorants : environ 100 personnes), la moitié tape *en particulier* dans la case de recherche en ligne, le plus souvent sans guillemet, et l'autre moitié adopte des stratégies variées, de type : recopier la question dans la case (ce qui est fort pertinent, mais non approprié à la recherche en ligne telle qu'elle est proposée actuellement) ou bien indiquer une partie de la question, souvent la fin (« expression en particulier »). Peu d'étudiants mettent des guillemets et aucun n'utilise la recherche avancée. Comme le précisent Boulton et Tyne (2014), les étudiants ont besoin de formation à la recherche en ligne, formation qui contribue, nous semble-t-il, à l'appropriation de la méthodologie de la recherche universitaire. L'appropriation de l'outil est donc progressive, allant d'outils que les apprenants connaissent bien ou partiellement, vers les corpus numériques souvent inconnus des étudiants. Ainsi, la 2^e partie de la consigne du dernier exercice invite l'enseignant à guider les apprenants vers ces corpus. En effet, la recherche sur le web permet difficilement d'extraire le type d'éléments que les corpus numériques nous offrent, sous forme de concordancier. Le premier corpus auquel ils ont été confrontés est Le corpus français de Leipzig⁹ parce qu'il propose des chiffres de fréquence d'utilisation immédiate. Une aide à la lecture des données est proposée sous forme d'échange collectif, puis, chacun cherche un mot de son choix et doit expliquer les données obtenues d'abord à son voisin, puis au groupe (Kübler, 2014). Ensuite, pour présenter un concordancier, nous commençons par Lextutor¹⁰ qui est un petit corpus avec lequel

⁸ URL : <https://books.google.com/ngrams>

⁹ URL : http://corpora.uni-leipzig.de/fr?corpusId=fra_mixed_2012

¹⁰ URL : <https://lxtutor.ca/conc/fr/>

l'étudiant peut lire et trier les extractions. Nous les encourageons alors à comparer les résultats des corpus comme dans l'exemple ci-dessous :



Copie d'écran 3 : Comparaison du lemme « considérer » dans Lextutor et Le Corpus Français de Leipzig

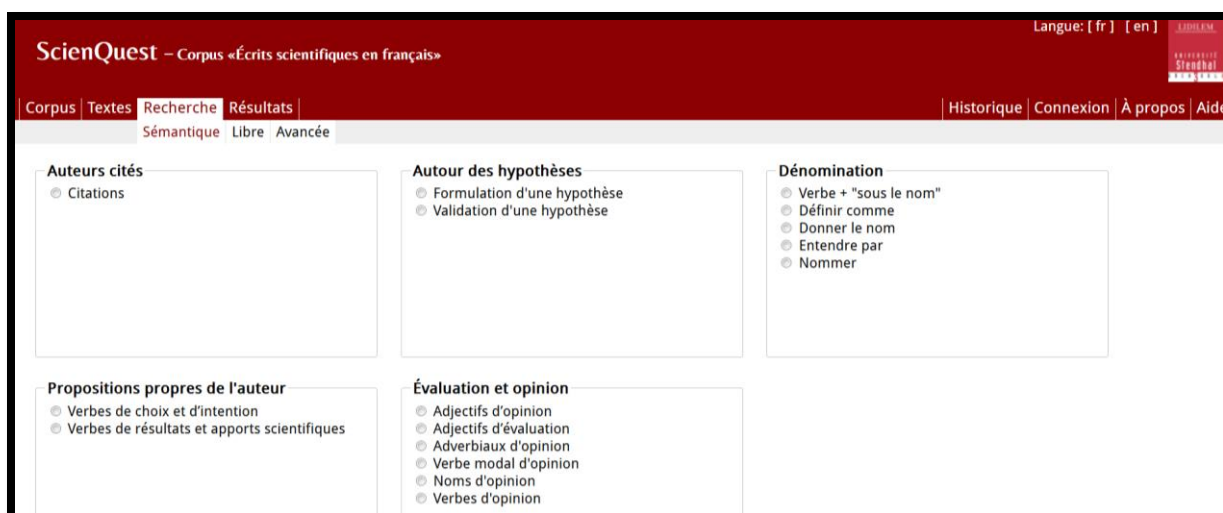
L'exemple de la copie d'écran 3 permet de voir que le verbe « considérer » est majoritairement suivi en français de « que » ou de « comme ». Les deux corpus le montrent que ce soit sous forme d'un concordancier (à droite celui de Lextutor) ou de chiffre de fréquence (à gauche dans le Corpus Français de Leipzig).

Enfin, nous leur présentons le Lexicoscope¹¹ dans lequel les extractions sont nombreuses et l'interface conviviale. Ils doivent alors chercher leurs premiers éléments phraséologiques et faire leurs premiers choix pour un tri spécifique, selon un objectif linguistique lié au choix de l'élément phraséologique, en accord avec l'enseignant. Souvent les étudiants commencent par des exemples cités en cours comme « émettre une hypothèse », puis chacun reprend des éléments trouvés dans des articles ou tentent des associations lexicales peu communes telles que « petits jeunes » ou « engendrer des résultats ».

Une fois les apprenants à l'aise avec les concordanciers de Lextutor, de Lexicoscope et les listes fréquentielles du Corpus Français de Leipzig, ils sont invités à découvrir ScienQuest puis le DiCorpus.

L'entrée sur ScienQuest suit la chronologie de la démarche d'interrogation proposée par le site : d'abord le choix du corpus, puis celui du type de recherche et enfin la requête (Tutin, 2013). Les étudiants sont invités à choisir – dans un premier temps – le corpus en français puis la recherche sémantique présentée ci-dessous :

¹¹ URL : <http://dip01.u-grenoble3.fr/~kraifo/lexicoscope/lexicoscope.php>



Copie d'écran 4 : Interface de la recherche sémantique dans ScienQuest

Cette entrée ne laisse pas de choix à l'apprenant, mais lui permet de visualiser rapidement ce que le site propose. L'utilisateur est invité à choisir une entrée sémantique (*l'opinion* par exemple) puis le type d'élément linguistique (les verbes), et un concordancier apparaît à ce propos avec des éléments pré-triés. Les étudiants sont alors invités à réfléchir aux choix des entrées sémantiques (notamment sur la place de ces éléments dans leurs écrits) pour en proposer d'autres. Ensuite, pour les guider vers la recherche libre, la réflexion s'oriente vers le détail des extractions : les variations au plan lexical (choix des verbes et des noms associés) et sémantique (divers sens apparaissent) sont à trier pour faire un choix personnel et utiliser l'élément qui paraît correspondre le mieux aux attentes du scripteur. Avant de passer à la recherche libre, DiCorpus est présenté. L'outil fournit des éléments pré-sélectionnés et donc oblige les apprenants à préciser leur questionnement linguistique pour leur écrit, notamment ce qu'il est important de mettre en avant dans leur rédaction. Ils doivent alors rédiger les questions qu'ils se posent et ensuite tenter de trouver les éléments appropriés à la rédaction scientifique pour énoncer leur pensée. Les questions récurrentes sont du type « Comment écrire mes questions de recherche ? » ou « Comment citer un auteur ? ».

Croiser ainsi les données des classes sémantiques proposées dans ScienQuest et celles autour des marqueurs discursifs de DiCorpus fait émerger à la fois des questions, mais aussi des prises de conscience de la part des apprenants. On constate leur soulagement quand ils s'aperçoivent que *par exemple* est très présent dans les écrits scientifiques et qu'ils peuvent l'utiliser sans complexe¹².

¹² Des représentations de l'écrit et de l'oral seraient à étudier dans ces contextes universitaires où les étudiants ont souvent de bon niveau en langue.

Le deuxième temps du cours est l'aide au choix des lemmes à chercher et donc l'aide à affiner les recherches sur les corpus. Ceci conduit à utiliser la recherche libre. C'est à partir de ce moment que les apprenants sont entraînés vers la démarche DDL (Data Driven Learning) puisqu'ils peuvent faire une requête sur n'importe quel lemme afin de découvrir simultanément le contenu du corpus et l'utilisation d'un lemme. Nous remarquons qu'ils ne sont pas complètement dans une démarche DDL, car le temps imparti pour un cours (parfois seulement 45mn selon les pays), ne permettent pas l'extraction et le tri de n-grammes pour une découverte exhaustive des trésors cachés du corpus.

Nous constatons que les étudiants-apprenants s'approprient rapidement l'interface. Ils associent de façon fructueuse les outils présentés en parallèle surtout le CNRTL¹³ qui donne accès à de multiples ressources autour de la langue. L'intérêt des étudiants pour les corpus croît et ils s'interrogent alors parfois sur le contenu d'autres interfaces qu'ils connaissent. Comme le soulignent Boulton et Tyne (2014), il est fréquent que les étudiants découvrent les corpus, mais il est tout aussi fréquent qu'ils les maîtrisent rapidement voire qu'ils en détournent l'utilisation pour l'adapter à leurs besoins pour leurs rédactions scientifiques. Qu'en est-il de leur appropriation des éléments phraséologiques ?

3. Découverte et appropriation des éléments phraséologiques

L'introduction des corpus dans la classe est rapidement suivie par l'introduction des éléments phraséologiques. Sans prononcer « élément phraséologique », les étudiants prennent conscience après quelques exemples, de la fréquence de ces formes figées ou semi-figées dans la langue. L'entrée dans ces éléments linguistiques passe par la traduction qu'ils ne trouvent pas littéralement¹⁴ et les oblige à envisager de nouvelles stratégies de mémorisation. Le chemin que les liens neuronaux du cerveau doivent parcourir est alors à envisager en relation directe avec le sens de l'élément phraséologique et non avec sa forme. C'est en ayant pour objectif l'aide à cette mémorisation voire à la mise en place de nouvelles stratégies d'apprentissage que nous avons développé les activités suivantes.

D'abord nous avons mis en relation la forme et le sens grâce aux entrées sémantiques de ScienQuest et DiCorpus. Le sens est premier et la forme suit immédiatement dès l'instant où

¹³ Centre National de ressources Textuelles et Lexicales : <http://www.cnrtl.fr/>

¹⁴ L'exemple de « poser une question » est clair pour tous car ils ont l'anglais comme L1, L2 ou L3 et connaissent « to ask a question ».

l'enseignant laisse les apprenants chercher seuls dans l'interface de requêtes libres de ScienQuest. Ainsi, les apprenants sont conduits à envisager les mots qu'ils cherchent pour énoncer leurs réflexions scientifiques, non pas à partir d'autres mots, mais à partir de concepts dans lesquels s'insèrent ces mots : *avancer une question* contient un verbe dit « d'opinion » et appartient, dans le classement de ScienQuest, à la classe sémantique « Évaluation et Opinion ». L'expérience montre que nombre d'étudiants n'a pas l'idée *a priori* d'utiliser le verbe *avancer* dans un tel cadre sémantique. Nous encourageons donc l'apprenant et l'enseignant à adopter une démarche onomasiologique, c'est-à-dire de se situer d'abord dans un concept, puis de consulter les entrées lexicales possibles et s'attarder sur l'une d'entre elles si nécessaire (approche sémasiologique). L'expérience montre que l'apprenant réagit d'abord de façon inverse (entrée sémasiologique) : il choisit un mot qu'il tente de traduire et de placer au bon endroit. Les outils présentés permettent l'inverse et surtout donnent à voir les mots possibles et les plus fréquents, c'est-à-dire ceux qui ne correspondent pas systématiquement à ceux envisagés par l'apprenant. On constate alors que les apprenants écartent la traduction après quelques exemples d'impossibilité. Ainsi, entrent-ils, de fait, dans la démarche onomasiologique qui simultanément les entraîne davantage encore dans une réflexion quant à leur rédaction.

Les observations des cours et des stratégies des étudiants ont permis d'aboutir à la séquence suivante : une fois l'utilisation des corpus maîtrisée ainsi que la compréhension du lien sémantique entre des classes sémantiques et les mots présents dans ces classes, alors les apprenants entrent facilement dans des exercices de repérage tels que présentés ci-après. L'exercice s'effectue à partir de l'extrait de l'introduction d'un article pris dans ScienQuest et des questions de découpage sémantique de l'extrait sont alors proposées aux étudiants, en petits groupes, pour une première approche. Voici l'extrait et les consignes afférentes :

« L'analyse linguistique que nous proposons s'inscrit en réponse à une demande de la SNCF relative à la perception du confort global en train. Les objectifs du projet sont d'identifier les propriétés sémantiques du confort en train et leurs relations de dépendance, à partir d'analyses linguistiques et cognitives. Cet article porte sur la manière dont les formes linguistiques en contexte, utilisant les ressources de la langue mises en discours, renseignent sur les structures cognitives construites à partir des perceptions sensorielles. On tente ici d'identifier, à partir de l'analyse des formes adjectivales, les représentations individuelles et partagées qui se construisent dans les discours des voyageurs, lorsqu'on les interroge sur leur expérience sensible du confort. <...> La première question que l'on se pose

alors est la suivante : le confort en train est-il une catégorie cognitive de ce type ?

(Scientext – article – Linguistique – Introduction)

1. D'après vous, cet extrait vient de quelle partie de l'article ? Pourquoi ?

A. résumé // B. introduction // C. théorie // D. méthodologie // E. conclusion

2. Découpez l'extrait en paragraphes et justifiez votre choix en précisant les mots et les thématiques de votre découpage : contexte, sujet, objectif, question.

3. Trouvez des expressions synonymes pour dire « on tente de... » à l'aide de l'outil CNTRL en ligne. »

La réponse attendue pour la distinction des parties est la suivante :

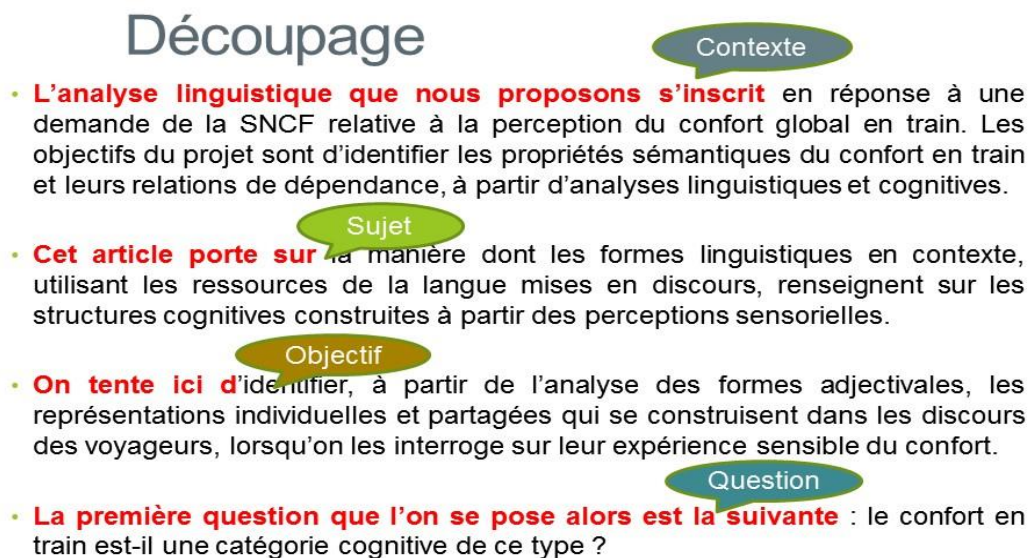


Figure 1 : Découpage d'un extrait de ScienQuest

Ce premier exercice permet d'introduire certaines routines verbales du lexique scientifique transdisciplinaire (ou LST de Tutin (2007)). Les exercices suivants sont alors plus précis, car ils sont développés pour s'attarder sur l'une des thématiques telles que « l'introduction du sujet » ou « les questions de recherche ». Après chaque repérage, les apprenants doivent recenser les nouvelles routines rencontrées et ainsi continuer d'alimenter leur LST. Par exemple, au fil des extraits et des repérages collectifs et individuels sur corpus, les apprenants peuvent retenir que pour introduire des questions de recherche, on peut utiliser : *il semble légitime de s'interroger sur, il est possible de formuler cette interrogation comme suit, Nous nous sommes demandé si...* De même pour présenter son sujet de recherche, on peut retenir : *Cet article porte sur, Dans cet article nous nous intéresserons à, Nous montrerons que...* Remarquons qu'il leur est demandé d'être attentifs à l'insertion syntagmatique de tels éléments et notamment ici en début de phrase (la majuscule l'atteste) ou non.

Une fois ces éléments rassemblés avec des informations morphosyntaxiques (l'insertion appropriée des éléments), lexicales (le choix des mots associés) et sémantiques (la place rhétorique de l'élément), nous demandons aux étudiants de rédiger une introduction (ou une conclusion ou un développement selon l'entrée choisie) à l'aide de tous les outils auxquels ils ont désormais accès.

Nous constatons un effet bénéfique de l'utilisation des corpus pour l'apprentissage des éléments phraséologiques. En effet, au fil des enseignements dispensés, nous constatons que les étudiants connaissent certains des éléments phraséologiques fréquents mais présentent des difficultés à les utiliser à bon escient. En outre, leur intuition linguistique ne les aide pas à savoir pourquoi insérer certains plutôt que d'autres. L'indication de fréquence trouvée dans les corpus les rassure quant au choix lexico-sémantique tandis que les exemples les aident pour l'insertion syntagmatique. Ainsi, pouvons-nous donner quelques résultats à partir des données recueillies auprès des étudiants de Master du niveau B2-C1.

4. Utilisation pour l'aide à la rédaction personnelle

L'appropriation a été vérifiée, pour les étudiants de Master de niveau B2-C1, à l'aide d'une présentation orale et écrite de leur sujet de recherche. Les éléments phraséologiques devaient à la fois figurer sur le diaporama¹⁵ de présentation et dans le discours oral l'accompagnant. Nous avons alors retrouvé des éléments

- du LST : *Cette recherche s'adresse à, Cet article porte sur, Nous montrerons que, il semble légitime de s'interroger sur, Nous pourrions donc avancer l'hypothèse que, nous nous sommes demandé si...*
- de DiCorpus : *En d'autres termes, en particulier, Dans un premier temps suivi de Dans un second temps.*

Ce qui est encourageant, c'est l'appropriation créative. Certains étudiants ont fait des associations qui sont tout à fait acceptables dans l'écrit scientifique, mais que nous n'avions pas vues en cours : *Notre première supposition est, Notre seconde supposition est.* Nous savons que certains les ont trouvées dans le corpus tandis que pour d'autres, nous n'avons pas d'indices quant à la stratégie développée. Si nous comparons avec des enseignants en formation continue, nous avons constaté qu'ils retrouvaient les extraits dans le corpus, qu'ils

¹⁵ Une courte séance a eu lieu pour l'aide à l'élaboration d'un diaporama.

comparaient alors avec d'autres extraits contenant un élément phraséologique proche sémantiquement et associaient cela à leur imagination. De fait, nous avons eu des résultats plus riches en termes linguistiques, mais il s'agissait d'enseignants de FLE.

Au fil des exercices, les apprenants étaient fortement encouragés à toujours chercher d'autres éléments dans les corpus. La pédagogie adoptée est celle de la collaboration. Tous les exercices étaient faits en groupe, parfois de petits défis intergroupes étaient proposés et chaque exercice devait finir par un récapitulatif personnel, sous forme de liste des éléments phraséologiques, de notes à propos d'une procédure pour trouver telle information dans tel corpus, etc. L'aide à la rédaction s'est donc sans cesse déroulée sur deux plans : un premier plan outillé avec les corpus et les sites en ligne, et un deuxième plan linguistique avec les éléments phraséologiques du LST. Nombre d'étudiants étaient d'abord impressionnés par l'interface de ScienQuest et se servaient plus aisément de Lextutor (dont l'interface n'est pourtant pas si aisée) et du Lexicoscope. Cependant, la majorité d'entre eux s'est finalement tournée vers ScienQuest en raison des requêtes infructueuses dans les corpus non spécialisés. Se dirigeront-ils davantage vers Lexicoscope dans lequel apparaissent désormais des textes scientifiques ? Il reste une étude à faire à ce propos.

5. Conclusion et discussion

Les séquences développées ont une entrée outillée accompagnée d'une démarche onomasiologique directe pour l'introduction des éléments phraséologiques du LST. Ceci s'est avéré fructueux dans les productions écrites et orales des étudiants.

Dans le cadre de la thèse de Tran (2014), les exercices, à partir d'extraits longs de ScienQuest, permettaient de faire des distinctions sémantiques fines entre des marqueurs discursifs. L'exemple de la distinction sémantique entre *à savoir* et *c'est-à-dire* dans des articles de recherche montre une complexité difficile à surmonter pour des apprenants non spécialistes de la langue. Ces exercices ont permis des distinctions sémantiques et la prise en compte de la ponctuation. Ce travail a été repris dans DiCorpus qui est apprécié des étudiants. Enfin, dans la thèse de Yan (2017), les facteurs d'aide à la mémorisation sont réunis. Les classes sémantiques permettent une entrée didactique plus abordable pour des apprenants non-linguistes, car elles représentent des notions qu'ils reconnaissent. Le fait de devoir comparer, décrire ou examiner un élément apparaît comme une action accessible tandis que la

« démarcation » reste un concept et donc un exercice difficile pour un étudiant débutant en écriture scientifique.

Ces différents essais ont abouti à la séquence présentée ci-dessus et ont permis aux étudiants d'aborder plus sereinement leur rédaction. L'entrée outillée guide les apprenants qui voient alors la langue comme un réel élément à manipuler. Ensuite, leur objectif rédactionnel les oblige à voir la langue dans toute sa complexité sémantique. Pour l'enseignant, l'utilisation des corpus devient l'objectif premier des séquences mais aussi un tremplin vers l'introduction d'éléments complexes de la langue pour ensuite entraîner les apprenants à s'approprier la méthode d'interrogation du corpus et les éléments phraséologiques du LST. L'apprenant comprend rapidement l'interface d'interrogation des corpus pour effectuer des requêtes. L'autonomie d'interrogation du corpus ScienQuest est facilitée par les entrées de la recherche sémantique immédiatement compréhensible par les apprenants.

Pour reprendre les questions soulevées au fil de l'article, nous avons constaté que les apprenants n'avaient plus systématiquement recours à la traduction dès l'instant où l'outil les aidait à voir les éléments et qu'ils avaient compris que la traduction littérale n'était que rarement possible. Notons que pour l'enseignement du lexique, même si les neurosciences commencent à nous donner des pistes sur le fonctionnement de la mémoire, la question de « l'aide à la compréhension des discours » reste floue (qu'est-ce que comprendre, au fond ?). La question de « comment apprendre à apprendre » trouve une réponse ici en abordant le contenu à acquérir par l'outillage. Nous espérons guider l'apprenant vers l'autonomie de l'utilisation des corpus et des éléments linguistiques à utiliser pour rédiger. En d'autres termes, une fois que l'apprenant sait où chercher l'information utile à sa rédaction personnelle, une fois qu'il a pris conscience de l'existence du LST, alors son autonomie d'apprentissage et de production scientifique ne peut être qu'en bonne voie. En outre, l'aspect affectif présent dans ce genre d'écrit, et notamment le stress de l'écriture, diminue au fil de la découverte de l'outil dans un premier temps et des éléments linguistiques appropriés dans un second temps. Nous envisageons désormais de mieux prendre en compte les affects dans ces cours pour l'aide à la rédaction scientifique.

Bibliographie

BARD, A., 2007, *Constitution d'un corpus d'écrits universitaires d'étudiants étrangers*. Université Stendhal-Grenoble3, Mémoire de Master.

- BOULTON, A. & TYNES, H., 2014, *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*, Paris, Didier.
- CAVALLA, C., 2014, "Collocations transdisciplinaires : réflexion pour l'enseignement", *Outils et méthode d'apprentissage en phraséodidactique*, in GONZALEZ-REY, M.I. (éds), Fernelmont, Belgique: E.M.E., 151-69.
- CAVALLA, C., 2015, "Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE", *Linx* 72, 95-110.
- CHAMBERS, A. & O'SULLIVAN, I., 2004, "Corpus consultation and advanced learners' writing skills in French", *ReCALL* 16-1, 158-72.
- HATIER, S., 2016, *Identification et analyse linguistique du lexique scientifique transdisciplinaire. Approche outillée sur un corpus d'articles de recherche en SHS*, Grenoble, Université Grenoble Alpes, Thèse de doctorat.
- KÜBLER, N., 2014, "Mettre en oeuvre la linguistique de corpus à l'université - Vers une compétence utile pour l'enseignement/apprentissage des langues ?", *Les Cahiers de l'Acedle* 11, 37-77.
- LECLERCQ, G., 2003, "Quelques usages de l'activité d'ingénierie de formation", *Savoirs* 2003, 71-104.
- MANGIANTE, J.-M. & PARPETTE, C., 2011, *Le français sur objectif universitaire*, Grenoble, Presses universitaires de Grenoble.
- MROUÉ, M., 2014, *Éléments pour une didactique du lexique transdisciplinaire dans les écrits de recherche en français langue étrangère*, Grenoble, Grenoble Alpes, Thèse de doctorat.
- TRAN, T. T. H., 2014, *Description de la phraséologie transdisciplinaire des écrits scientifiques et réflexions didactiques pour l'enseignement à des étudiants non-natifs. Application aux marqueurs discursifs*, Grenoble, Grenoble Alpes, Thèse de doctorat.
- TRAN, T. T. H. & FALAISE, A., 2016, "Quelles stratégies pédagogiques pour une introduction des corpus en classe de langues ?", *Colloque Langues sur objectifs spécifiques : perspectives croisées entre linguistique et didactique*, 24-25 nov, Université Grenoble Alpes.
- TUTIN, A., 2007, "Autour du lexique et de la phraséologie des écrits scientifiques", *Revue française de linguistique appliquée* XII-2, 5-13.
- TUTIN, A., 2013, "La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétoriques", *L'écrit scientifique : du lexique au discours. Autour de Scientext*, in TUTIN, A. & GROSSMANN, F. (éds), Rennes: Presses Universitaires de Rennes, 27-43.
- YAN, R., 2017, *Étude des constructions verbales scientifiques dans une perspective didactique : utilisation des corpus dans le diagnostic des besoins langagiers du fle à l'aide des techniques de tal*, Grenoble, Université Grenoble Alpes, Thèse de doctorat.